EMPIRICAL SAMPLING STUDY OF A GOODNESS OF FIT STATISTIC FOR DENSITY FUNCTION ESTIMATION

P. A. W. Lewis, et al

Naval Postgraduate School

Prepared for:

National Science Foundation

March 1975

NPS55Lw75031

NAVAL POSTGRADUATE SCHOOL
Monterey, California

EMPIRICAL SAMPLING STUDY OF A GOODNESS OF FIT STATISTIC FOR DENSITY FUNCTION ESTIMATION

by

P. A. W. Lewis, L. H. Liu,

D. W. Robinson and M. Rosenblatt

March 1975

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPS55Lw75031	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Empirical Sampling Study of a Goodness of Fit Statistic for Density Function Estimation		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) P. A. W. Lewis, L. H. Liu, D. W. Robinson and M. Rosenblatt		8. CONTRACT OR GRANT NUMBER(s) AG-476
9. PERFORMING ORGANIZATION NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE March 1975 13. NUMBER OF PAGES 30
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Goodness of fit statistic
Density function estimation
Empirical sampling
Histogram estimates
Kolmogorov-Smirnov test

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The distribution of a measure of the distance between a probability density function and its estimate is examined through empirical sampling methods. The estimate of the density function is that proposed by Rosenblatt using sums of weight functions centered at the observed values of the random variables. The weight function in all cases was triangular, but both uniform and Cauchy densities were tried for different

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

PRICES SUBJECT TO CHANGE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Section 20 continued.

sample sizes and bandwidths. The simulated distributions look as if they could be approximated by Gamma distributions, in many cases. Some assessment can also be made of the rate of convergence of the moments and the distribution of the measure to the limiting moments and distribution, respectively.

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral Isham Linder
Superintendent

Jack R. Borsting
Provost

The work reported herein was supported in part by the Office of Naval Research and the National Science Foundation, AG-476.

Prepared by:

Peter A.W. Lewis
Peter A. W. Lewis, Professor
Department of Operations Research
and Administrative Sciences

Peter A.W. Lewis for L.H. Liu
L. H. Liu

D W. Robinson by P. Lewis
D. W. Robinson

M Rosenblatt by P. Lewis
M. Rosenblatt

Reviewed by:

David A. Schrady
David A. Schrady, Professor
Department of Operations Research
and Administrative Sciences

Released by:

Robert Fossum
Robert Fossum
Dean of Research

1. INTRODUCTION

There are several recently proposed classes of empirical probability density function [1,4,5,7] all generally considered to be superior to the classical histogram estimates. The class considered in this paper is based on independent observations, i.e. $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with continuous unknown density function $f(x)$. The method used to estimate $f(x)$ is that proposed by Rosenblatt; denoting the estimate by $f_n(x)$, we define

$$f_n(x) = \frac{1}{nb(n)} \sum_{j=1}^{n} W\left[\frac{x - X_j}{b(n)}\right],$$

where $W(u)$ is a bounded non-negative integrable weight function with

$$\int_{-\infty}^{\infty} W(u)\,du = 1,$$

and $b(n)$ is a positive bandwidth function which tends to zero as $n \to \infty$, but is such that $o[b(n)] = 1/n$. Thus we might have $b(n) \sim n^{-1/2}$, for example.

We note that all estimates of this form are themselves density functions for a given set of observations; that is,

$$f_n(x) \geq 0,$$

$$\int_{-\infty}^{\infty} f_n(x)\,dx = 1 .$$

Since the X_j's are random variables, $f_n(x)$ is a continuous parameter stochastic process, but it is clearly non-stationary.

The estimate $f_n(x)$ can be shown to be locally biased for any value of x under relatively mild conditions [4]. Our object in this paper is to investigate a global measure of how good $f_n(x)$ is as an estimate of $f(x)$. The measure was originally proposed by Bickel and Rosenblatt [2] and is given by

$$\beta(n) = \int \frac{[f_n(x) - f(x)]^2}{f(x)} dx .$$

Since the value of $\beta(n)$ will vary with each realization of $X_1,\ldots,X_n$, it is a statistic or function of the n random variables. A possible application for such a statistic would be in goodness-of-fit type tests, in an analogous manner to the more familiar Kolmogorov-Smirnov test.

Bickel and Rosenblatt [2] have established that if $b(n) = o[n^{-2/9}]$ as $n \to \infty$ and if $a(x)$ is a bounded, piecewise smooth integrable function then

$$b(n)^{-1/2}[nb(n)\int[f_n(x)-f(x)]^2 a(x)\,dx - \int f(x)a(x)\,dx \int W(z)^2 dz]$$

is asymptotically normally distributed with zero mean and variance

$$2W^{(4)}(0)\int a(x)^2 f(x)^2 dx ,$$

as $n \to \infty$, where $W^{(4)}(0)$ is the fourth convolution of W with itself. Thus, $\beta(n)$ has an asymptotically normal distribution, regardless of the underlying density $f(x)$.

A problem in this situation is that, unlike the Kolmogorov-Smirnov test statistic, the statistic $\beta(n)$ is

not distribution-free. Further, its exact distribution for any finite value of n does not seem to be mathematically tractable. We thus examined some representative cases through simulation, hoping that $\beta(n)$ would be fairly robust with rapid convergence to the asymptotic distribution. It was also hoped that the simulations would cast light on these conjectures and perhaps suggest some unexpected results.

2. SIMULATION

The primary object of the simulation was to investigate the distribution of the statistic $\beta(n)$:

$$\beta(n) = \int \frac{[f_n(x) - f(x)]^2}{f(x)} dx ,$$

over a suitable range of integration. We performed simulations with synthetic sampling from both uniform and Cauchy distributions; the triangular weight function

$$W(u) = \begin{cases} 1 - |u|, & \text{if } |u| \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

was used to evaluate $f_n(x)$ in both cases. We found little difference as far as $\beta(n)$ was concerned between the triangular and other "smoother" (e.g., quadratic) weight functions for our samples of from 100 to 1500 deviates.

A. UNIFORM RANDOM VARIABLES

In the case of uniform (0,1) random variables, we have

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} . \end{cases}$$

Thus, $\beta(n)$ becomes,

$$\beta(n) = \int_{b(n)}^{1-b(n)} [f_n(x) - 1]^2 dx . \qquad (2.1)$$

The limits of integration are from $b(n)$ to $1 - b(n)$ instead of from 0 to 1 to avoid the marked bias of $f_n(x)$ near 0 and 1. As long as $b(n) \leq x \leq 1-b(n)$, though, $f_n(x)$

is unbiased:

$$E[f_n(x)] = \frac{1}{b(n)}\int_0^1 W\left[\frac{x-y}{b(n)}\right] dy$$

$$= \frac{1}{b(n)}\int_{x-b(n)}^{x+b(n)} \left[1 - \frac{|x-y|}{b(n)}\right] dy$$

$$= \frac{1}{b(n)}\left[\int_{x-b(n)}^{x+b(n)} dy - \int_{x-b(n)}^{x} \frac{x-y}{b(n)}\, dy - \int_{x}^{x+b(n)} \frac{y-x}{b(n)}\, dy\right]$$

$$= \frac{1}{b(n)}\left[2b(n) - \frac{1}{b(n)}\left[\frac{b(n)}{2}\right]^2 - \frac{1}{b(n)}\left[\frac{b(n)}{2}\right]^2\right]$$

$$= 1 .$$

Also, for the same range of x,

$$Var[f_n(x)] = Var\left[\frac{1}{nb(n)}\sum_{j=1}^{n} W\left[\frac{x - X_j}{b(n)}\right]\right]$$

$$= \frac{1}{n^2 b(n)^2}\sum_{j=1}^{n} Var\, W\left[\frac{x - X_j}{b(n)}\right]$$

$$= \frac{1}{nb(n)^2} Var\, W\left[\frac{x - X_j}{b(n)}\right]$$

$$= \frac{1}{nb(n)^2}\left[\int_0^1 W^2\left[\frac{x-y}{b(n)}\right] dy - \left[\int_0^1 W\left[\frac{x-y}{b(n)}\right] dy\right]^2\right] .$$

Since $f_n(x)$ is a piecewise linear function when a triangular weight function is used, the integral in (2.1) can be evaluated in principle but the work becomes prohibitive for even moderate sample sizes. We thus approximated the integral using Simpson's rule with 100 equal subintervals. The results were found to be satisfactory in the sense that the value did not change appreciably when a finer grid (up to 500 subintervals) was used. In general, we found that a larger sample size required a finer grid; apparently the value of $f_n(x)$ changes

more rapidly over a small interval when n is large.

We used three different bandwidths in the uniform case: $3 / n^{1/2}$, $1 / n^{1/2}$ and $1 / n$. For each bandwidth sample sizes of 100, 200, 500, 1000 and 1500 were investigated so that a total of 15 experiments were carried out. Each experiment consisted of 2000 independent replications each of which resulted in the calculation of a single value of $\beta(n)$ using (2.1). The replications for a given experiment were divided into five sections of 400 observations each so that variability of the simulation results could be assessed between sections.

Besides the 400 observed values of $\beta(n)$, the computer output for each of the 75 sections included a histogram, an empirical log-survivor function plot, an empirical CDF plot and a normal probability plot. A histogram and an empirical log-survivor plot were also computed for the pooled sample of 2000 for each experiment. These plots are all reproduced in reference [3]; some of the more interesting cases are included in Section 4.

It was found that a better picture of the distribution of the data resulted when the empirical density function of the $\beta(n)$'s was plotted over the histogram plot. A fairly wide bandwidth was needed to suppress large fluctuations in $f_n(x)$; it was found that $b(n) = R / n^{1/2}$ was a fairly robust choice. (R denotes the sample range [maximum value - minimum value] of the $\beta(n)$ sample.) The solid lines in the Figures in Section 4 are empirical density estimates using this bandwidth and the triangular weight function.

B. CAUCHY RANDOM VARIABLES

The Cauchy density function is

$$f(x) = \frac{1}{\pi(1+x^2)} .$$

We used the same density estimator as in the uniform case:

$$f_n(x) = \frac{1}{nb(n)} \sum_{j=1}^{n} W\left[\frac{x - X_j}{b(n)}\right] ,$$

and again the triangular weight function. We chose a range of integration (-3,+3):

$$\beta(n) = \int_{-3}^{+3} \frac{[f_n(x) - f(x)]^2}{f(x)} dx .$$

This range comprises 80% of the probability mass for this distribution. Again, Simpson's rule was used to approximate the integral; in this case a grid of 600 subintervals was selected after examining 100, 300, 600 and 900 subinterval grids.

The Cauchy distribution was chosen because for finite n $f_n(x)$ has a bias component; this component usually decreases with bandwidth for a fixed value of n, although the pointwise variance of $f_n(x)$ increases with decreasing bandwidth. It seems likely that the variance of $\beta(n)$ would also decrease under these conditions, as indeed it was observed to do.

Three bandwidths were also employed in the Cauchy case: $1 / n^{1/2}$, $3 / n^{1/2}$ and $20 / n^{1/2}$, the last one representing a case in which bias in the estimator $f_n(x)$ plays a major

role in the distribution of $\beta(n)$. The same five sample sizes were used here for each bandwidth as were used for the uniform simulations; output from the fifteen Cauchy experiments was obtained just as in the uniform case.

3. TABULAR RESULTS AND GAMMA FITS

Using the asymptotic result obtained by Bickel and Rosenblatt [5], for a uniform random variable the quantity

$$b(n)^{-1/2}\{nb(n)\int_{b(n)}^{1-b(n)}|f_n(x)-1|^2dx - [1-2b(n)]\int W(u)^2du\}$$

is asymptotically normally distributed with mean 0 and variance

$$2W^{(4)}(0)\ [1-2b(n)]$$

as $n \to \infty$ if $nb(n) \to \infty$ and $b(n) = o(n^{-2/9})$. For the triangular weight function,

$$\int W(u)^2du = \frac{2}{3}$$

and $W^{(4)}(0)$, the fourth convolution of W with itself at zero, is 302/630.

From the above expressions, we get

$$E[\ \beta(n)\ ] = E\left[\int_{b(n)}^{1-b(n)}|f_n(x)-1|^2dx\right] \sim \frac{2}{3}\frac{1-2b(n)}{nb(n)}$$

$$Var[\ \beta(n)\ ] = Var\left[\int_{b(n)}^{1-b(n)}|f_n(x)-1|^2dx\right] \sim \frac{2W^{(4)}(0)[1-2b(n)]}{n^2b(n)}$$

Comparisons of the simulated values for the uniform experiments with the conjectured ones are tabulated in Table III.1 (means) and Table III.2 (variances). Especially for small bandwidth the agreement between the asymptotic and simulated variances is very good even for small n (n = 100). The same is true for expected value, although convergence is slower than for the variance and again slower for large bandwidth.

TABLE III.1 Comparison of estimated mean values and asymptotic mean values of $\beta(n)$ for different bandwidths and sample sizes.

n	$b(n)=3/\sqrt{n}$	$E(\beta(n))$	$E(\beta(n))/(1-2b(n))$	
			Conjectured	Computer output
100	.3000	.0089	.0222	.0127
200	.2121	.0090	.0157	.0109
500	.1342	.0073	.0099	.0075
1000	.0949	.0057	.0070	.0058
1500	.0775	.0048	.0057	.0051
	$b(n)=1/\sqrt{n}$			
100	.1000	.0533	.0667	.0583
200	.0707	.0405	.0471	.0415
500	.0447	.0271	.0298	.0269
1000	.0316	.0197	.0211	.0197
1500	.0258	.0163	.0172	.0168

TABLE III.2 Comparison of estimated standard deviation values and asymptotic standard deviation values of $\beta(n)$ for different bandwidths and sample sizes.

n	$b(n)=3/\sqrt{n}$	$\sigma(\beta(n))$	$\sigma(\beta(n))/(1-2b(n))$	
			Conjectured	Computer output
100	.3000	.0113	.0283	.0115
200	.2121	.0081	.0141	.0088
500	.1342	.0046	.0063	.0047
1000	.0949	.0029	.0036	.0030
1500	.0775	.0022	.0026	.0023
	$b(n)=1/\sqrt{n}$			
100	.1000	.0277	.0346	.0315
200	.0707	.0171	.0199	.0189
500	.0447	.0088	.0097	.0092
1000	.0316	.0053	.0057	.0056
1500	.0258	.0040	.0042	.0043

In contrast to the moments, the distribution of $\beta(n)$ converges very slowly. The complete results (reference [3]) reveal that the histograms and empirical density functions of the $\beta(n)$'s are all skewed to the right; see Figures IV.1 to IV.9 for examples.

The form of the histograms as well as the log-survivor plots suggested that the $\beta(n)$ statistic is approximately Gamma(θ,k) distributed, where the Gamma density is given by

$$f(x;\, k,\theta) = \frac{(x/\theta)^{k-1}\, e^{-x/\theta}}{\theta\,(k-1)!} ,$$

and the mean and variance are

$$E[X] = k\theta;$$
$$Var[X] = k\theta^2 .$$

Accordingly, estimates $\bar{K}$ and $\bar{\theta}$ of k and θ for each experiment were obtained from the sample of 2000 $\beta(n)$'s. Shenton and Bowman's almost unbiased estimators for the Gamma distribution [6] were used; these give reasonable results when $k \geq 0.5$, as in this case. The estimate values are tabulated in Table III.3; also tabulated are estimates of the standard deviation of $\bar{K}$ and $\bar{\theta}$ which were obtained from the five sections in each experiment. A parametric density estimate is thus obtained for the $\beta(n)$ sample; it may be compared with the non-parametric estimate $f_n(x)$ by examining the graphs in Section 4, where the Gamma density function is plotted with a dashed line.

TABLE III.3 Estimated Parameters for Fitted Gamma Distribution for $\beta(n)$.

DISTRIBUTION	b(n)	n	K	B
UNIFORM	$1 / \sqrt{n}$	100	3.969 ± 0.206	0.01390 ± .00095
		200	5.780 ± 0.659	0.00715 ± .00095
		500	8.881 ± 0.889	0.00311 ± .00029
		1000	13.011 ± 0.796	0.00153 ± .00008
		1500	17.316 ± 1.467	0.00095 ± .00008
	$3 / \sqrt{n}$	100	1.153 ± 0.048	0.00967 ± .00058
		200	1.718 ± 0.174	0.00588 ± .00078
		500	2.707 ± 0.241	0.00281 ± 0.00026
		1000	4.028 ± 0.241	0.00145 ± .00007
		1500	5.248 ± 0.423	0.00096 ± .00008
	$1 / n$	100	40.337 ± 2.555	0.01616 ± .00117
		200	39.511 ± 2.347	0.01675 ± .00106
		500	33.649 ± 1.820	0.01958 ± .00111
		1000	32.033 ± 3.305	0.02059 ± .00244
		1500	31.712 ± 1.999	0.02088 ± .00124
	$1 / \sqrt{n}$	100	22.362 ± 1.488	0.01745 ± .00114
		200	32.305 ± 2.022	0.00864 ± .00054
		500	60.147 ± 4.009	0.00293 ± .00022
		1000	79.897 ± 6.608	0.00157 ± .00014
		1500	101.100 ± 7.783	0.00102 ± .00007
CAUCHY	$3 / \sqrt{n}$	100	9.272 ± 0.406	0.01331 ± .00062
		200	12.744 ± 0.645	0.00709 ± .00037
		500	20.701 ± 1.673	0.00277 ± .00022
		1000	29.303 ± 2.541	0.00140 ± .00012
		1500	34.265 ± 3.963	0.00099 ± .00010
	$20 / \sqrt{n}$	100	7.103 ± 0.217	0.00776 ± .00035
		200	4.144 ± 0.069	0.00619 ± .00009
		500	3.445 ± 0.161	0.00312 ± .00016
		1000	4.211 ± 0.357	0.00152 ± .00009
		1500	5.385 ± 0.335	0.00095 ± .00005

4. GRAPHICAL RESULTS AND GENERAL DISCUSSION

The graphs for the following experiments have been reproduced from [3] because they give the greatest insight into the distribution of $\beta(n)$; these graphical results are more informative than the tabulated means, variances and Gamma fits of the previous Section.

Figure	Random Variable	n	b(n)	K
4.1	Uniform	200	$3/n^{1/2}$	1.718
4.2	Uniform	500	$1/n^{1/2}$	8.881
4.3	Uniform	1500	$1/n^{1/2}$	17.316
4.4	Uniform	200	$1/n$	39.511
4.5	Cauchy	100	$1/n^{1/2}$	22.362
4.6	Cauchy	100	$3/n^{1/2}$	9.272
4.7	Cauchy	1500	$20/n^{1/2}$	5.385
4.8	Uniform	1500	$3/n^{1/2}$	5.248
4.9	Uniform	100	$1/n^{1/2}$	3.969

In interpreting the graphs we can be guided by crude heuristics. In the case of a density estimate $f_n(x)$ with bandwidth b(n) there is dependence within a range of order b(n) and an approach to independence for points separated by a distance of order larger than b(n). Thus in the case of uniform random variables the integral $\beta(n)$ could be thought of as having the equivalent of the order of $[1-2b(n)]/b(n)$ independent summands. In the first case (Figure 4.1; n=200, $b(n)=3/\sqrt{n}$, K=1.718) we obtain

$$(1 - 3\sqrt{2}/10) \; / \; [3/(10\sqrt{2})] = 2.71 \; .$$

This is rather small so that one does not expect a good

Gaussian fit. We give $\bar{K}$ from the previous Section since $2\bar{K}$ may be interpreted as an equivalent number of degrees of freedom; the larger the fitted $\bar{K}$, the closer we are to normality. In a loose sense it is clear that a gamma fit is likely to be more appropriate and this is confirmed by looking at the graphs.

In the second case (Figure 4.2; $n=500$, $b(n)=1/\sqrt{n}$, $\bar{K}=8.881$) we have

$$(1 - \sqrt{2}/10)\ 10\sqrt{2} = 12.14 ,$$

which is a bit larger. It is interesting to note that the estimated (smoothed) density function of $\beta(n)$ gives us greater insight apparently in all cases. Here we see the beginning of an approach to asymptotic normality though it is still suggested that a Gamma fit might be appropriate. The next case (Figure 4.3; $n=1500$, $b(n)=1/\sqrt{n}$, $\bar{K}=17.316$) with

$$[\ 1 - 1/(5\sqrt{15})\ ]\ 10\sqrt{15} = 36.73$$

shows a closer approach to normality. It may be seen that the major departure between the parametric and non-parametric density estimates occurs in the vicinity of the mode where $f_n(x)$ tends to fluctuate about the true value. The fit in the tails appears excellent in all cases.

The next uniform case (Figure 4.4; $n=200$, $b(n)=1/n$, $\bar{K}=39.511$) is strictly speaking outside the range of results suggested by the paper of Bickel and Rosenblatt [2]. Here $f_n(x)$ is asymptotically compound Poisson rather than

asymptotically normal. Nonetheless we notice that it looks as if a Gaussian fit would be very good and this is consistent with the magnitude of our crude index

$$(1 - .01)\ 200 = 198 .$$

It would be interesting for someone to prove the suggested asymptotic normality.

In the simulation of sampling from a uniform distribution, the density estimator has no bias. To investigate the effect of bias, we repeated the uniform experiments for Cauchy-distributed random variables, integrated over the range $-3+b(n)$ to $3-b(n)$. The first case (Figure 4.5; $n=100$, $b(n)=1/\sqrt{n}$, $\bar{K}=22.362$) has index

$$(6 - .2)\ 10 = 58$$

and one notices that a Gaussian fit looks very good. The next case (Figure 4.6; $n=100$, $b(n)=3/\sqrt{n}$, $\bar{K}=9.272$) has index

$$(6 - .6)\ 10/3 = 17.66 ,$$

and a Gaussian fit looks fair but not good. In the last Cauchy case one expects substantial bias (Figure 4.7; $n=1500$, $b(n)=20/\sqrt{n}$, $\bar{K}=5.385$) and the crude index is

$$(6 - 4/\sqrt{15})\ \sqrt{15}/2 = 9.62 .$$

A Gamma fit is suggested. Altogether the effects of bias don't seem to be that extreme when sampling from the Cauchy distribution but this may be due to the fact that the Cauchy density is a very smooth function.

The last two cases involve sampling from the uniform distribution again but with different sample sizes and bandwidths. Figure 4.8 is for $n=1500$, $b(n)=3/\sqrt{n}$ and $\bar{K}=5.248$, while Figure 4.9 is for $n=100$ and $b(n)=1/\sqrt{n}$ for which $\bar{K}=3.969$.

The problem in using $\beta(n)$ as a measure of goodness of fit in the non-limiting Gamma case is to determine k and θ. If one wishes to fit the Gamma distribution using the method of moments, one can use the fact that the mean and variance of $\beta(n)$ should be approximately (on asymptotic grounds) $W^{(2)}(0)$ and $b(n)W^{(4)}(0)$, respectively. One might then use

$$\overset{*}{k} = \frac{[W^{(2)}(0)]^2}{b(n)W^{(4)}(0)},$$

$$\overset{*}{\theta} = \frac{\overset{*}{k}}{W^{(2)}(0)}$$

as estimates of k and θ. The results in Section 3 suggest that this procedure should produce adequate results except when there is appreciable bias in the density function estimate.

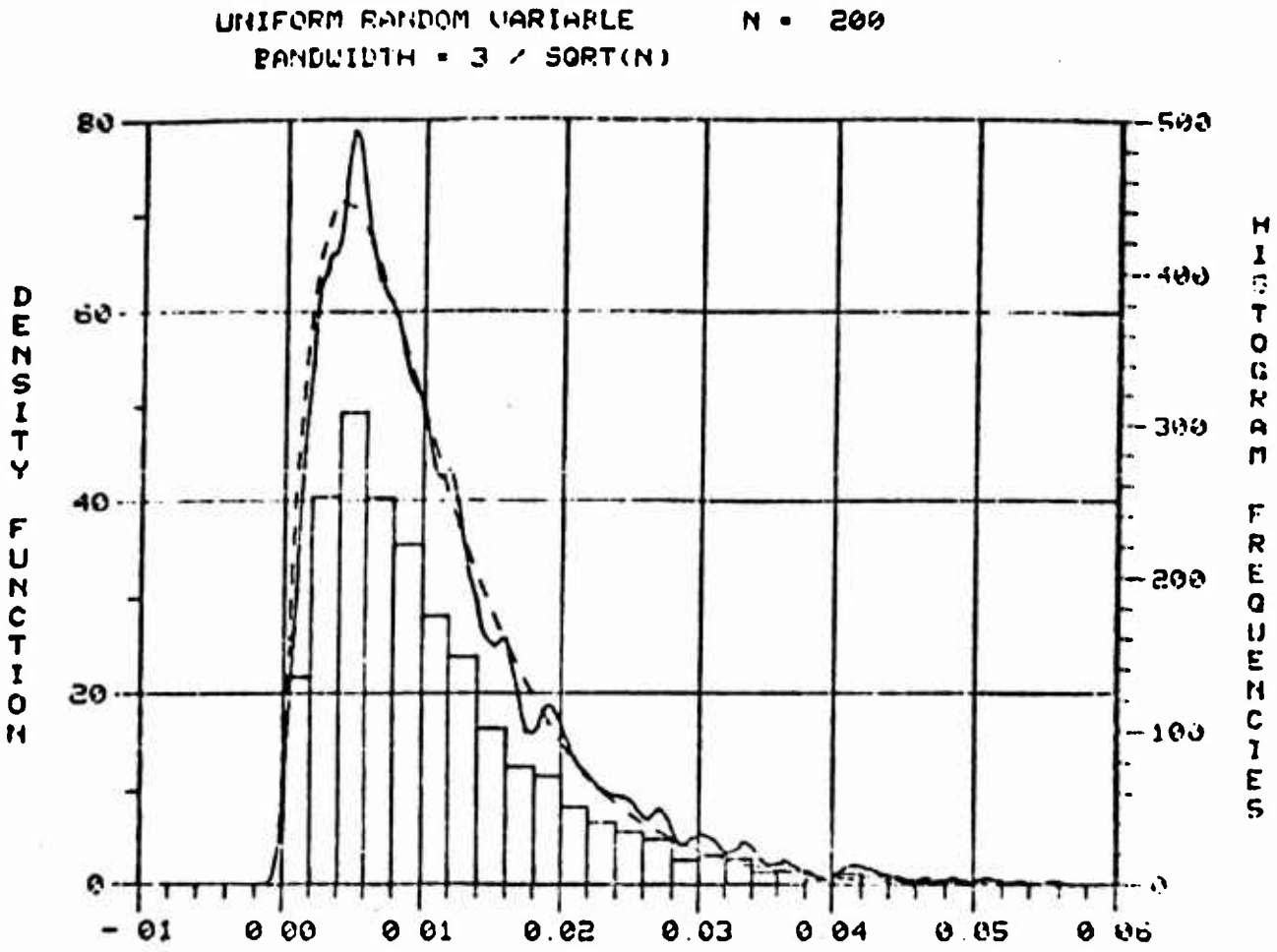

Figure 4.1. Distribution of the statistic $\beta(n)$ for a uniform random variable with n = 200 and bandwidth $3 / \sqrt{n}$. The solid line shows the Rosenblatt empirical density function of the $\beta(n)$'s while the dashed line is a fitted Gamma density function with $\bar{K} = 1.718$ and $\bar{\theta} = .00588$.

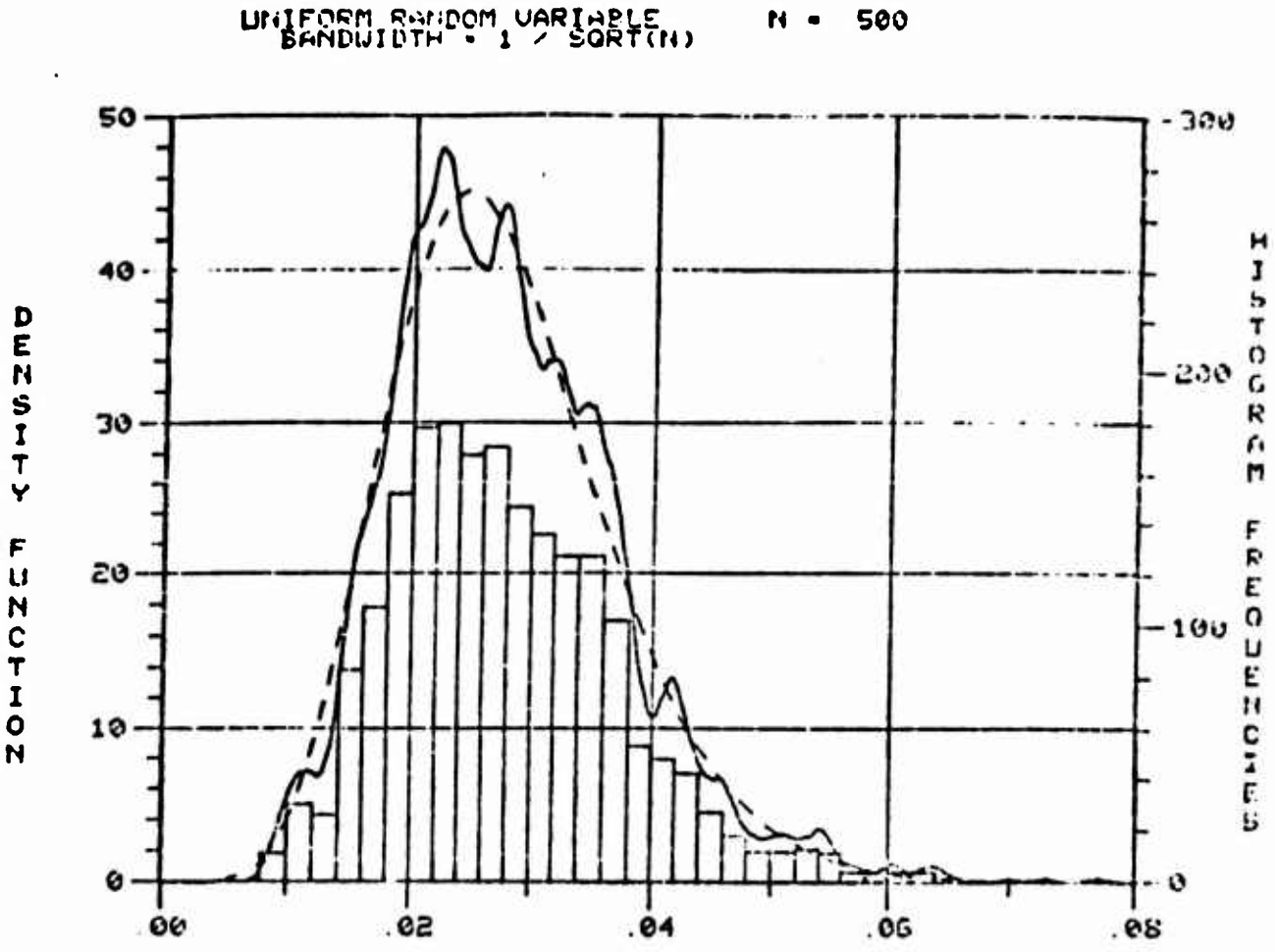

Figure 4.2. Distribution of the statistic $\beta(n)$ for a uniform random variable with n = 500 and bandwidth $1 / \sqrt{n}$. The solid line shows the Rosenblatt empirical density function of the $\beta(n)$'s while the dashed line is a fitted Gamma density function with $\bar{K} = 8.881$ and $\bar{\theta} = .00311$.

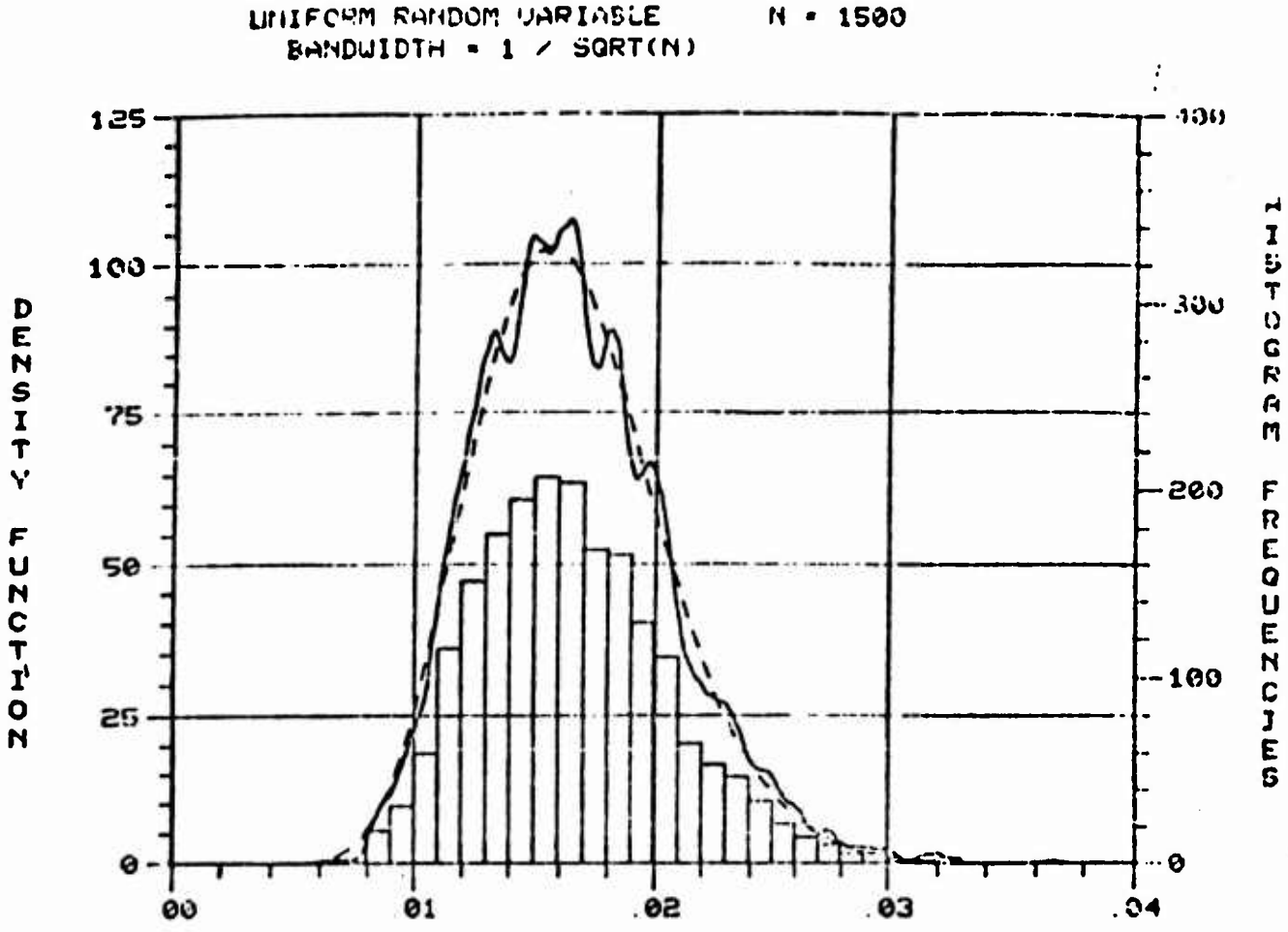

Figure 4.3. Distribution of the statistic $\beta(n)$ for a uniform random variable with n = 1500 and bandwidth $1 / \sqrt{n}$. The solid line shows the Rosenblatt empirical density function of the $\beta(n)$'s while the dashed line is a fitted Gamma density function with $\bar{K} = 17.316$ and $\bar{\theta} = .00095$.

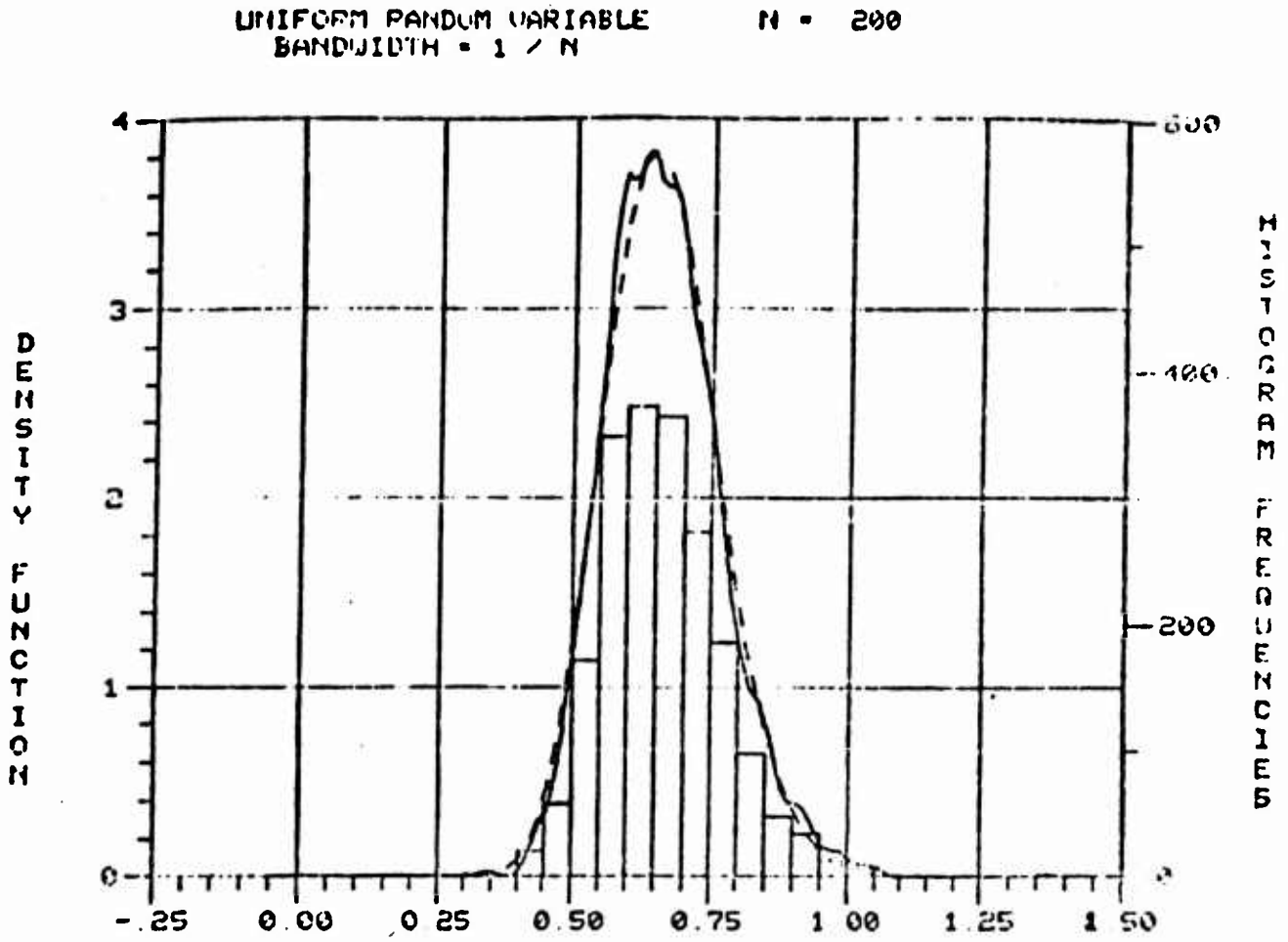

Figure 4.4. Distribution of the statistic $\beta(n)$ for a uniform random variable with n = 200 and bandwidth 1 / n. The solid line shows the Rosenblatt empirical density function of the $\beta(n)$'s while the dashed line is a fitted Gamma density function with $\bar{K}$ = 39.511 and $\bar{\theta}$ = .01675.

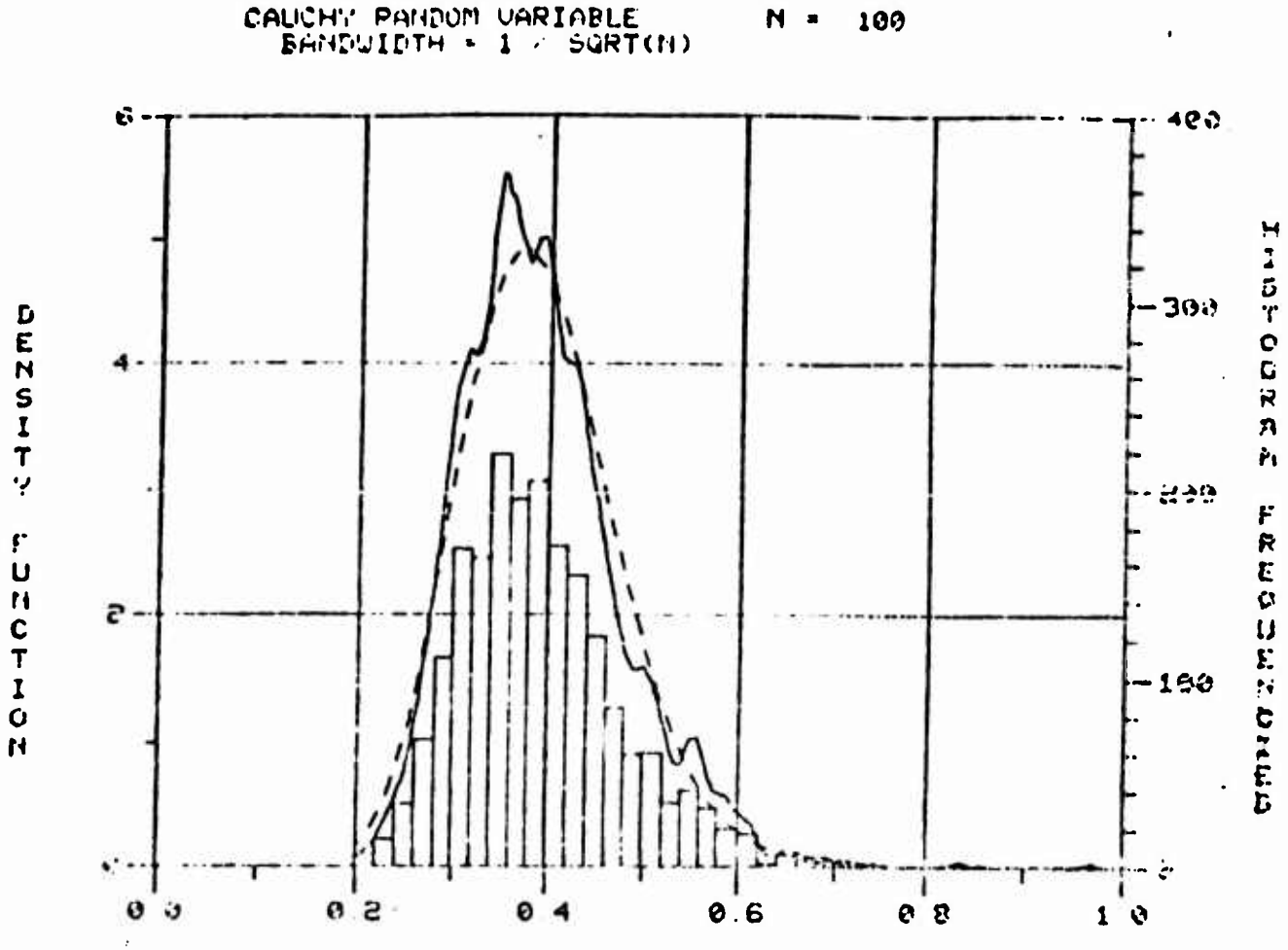

Figure 4.5. Distribution of the statistic $\beta(n)$ for a Cauchy random variable with n = 100 and bandwidth $1 / \sqrt{n}$. The solid line shows the Rosenblatt empirical density function of the $\beta(n)$'s while the dashed line is a fitted Gamma density function with $\bar{K} = 22.362$ and $\bar{\theta} = .01745$.

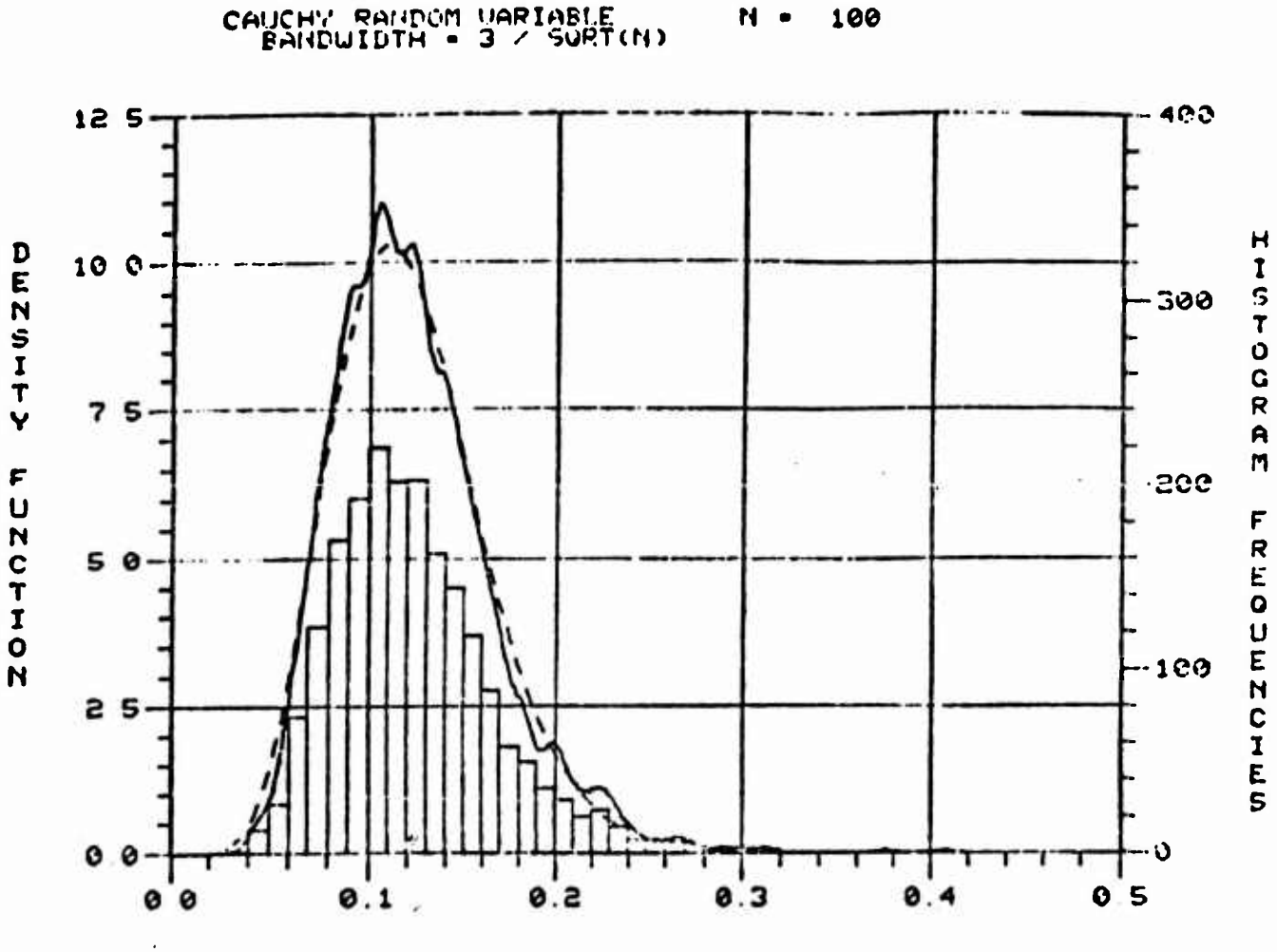

Figure 4.6. Distribution of the statistic $\beta(n)$ for a Cauchy random variable with n = 100 and bandwidth $3 / \sqrt{n}$. The solid line shows the Rosenblatt empirical density function of the $\beta(n)$'s while the dashed line is a fitted Gamma density function with $K = 9.272$ and $\theta = .01331$.

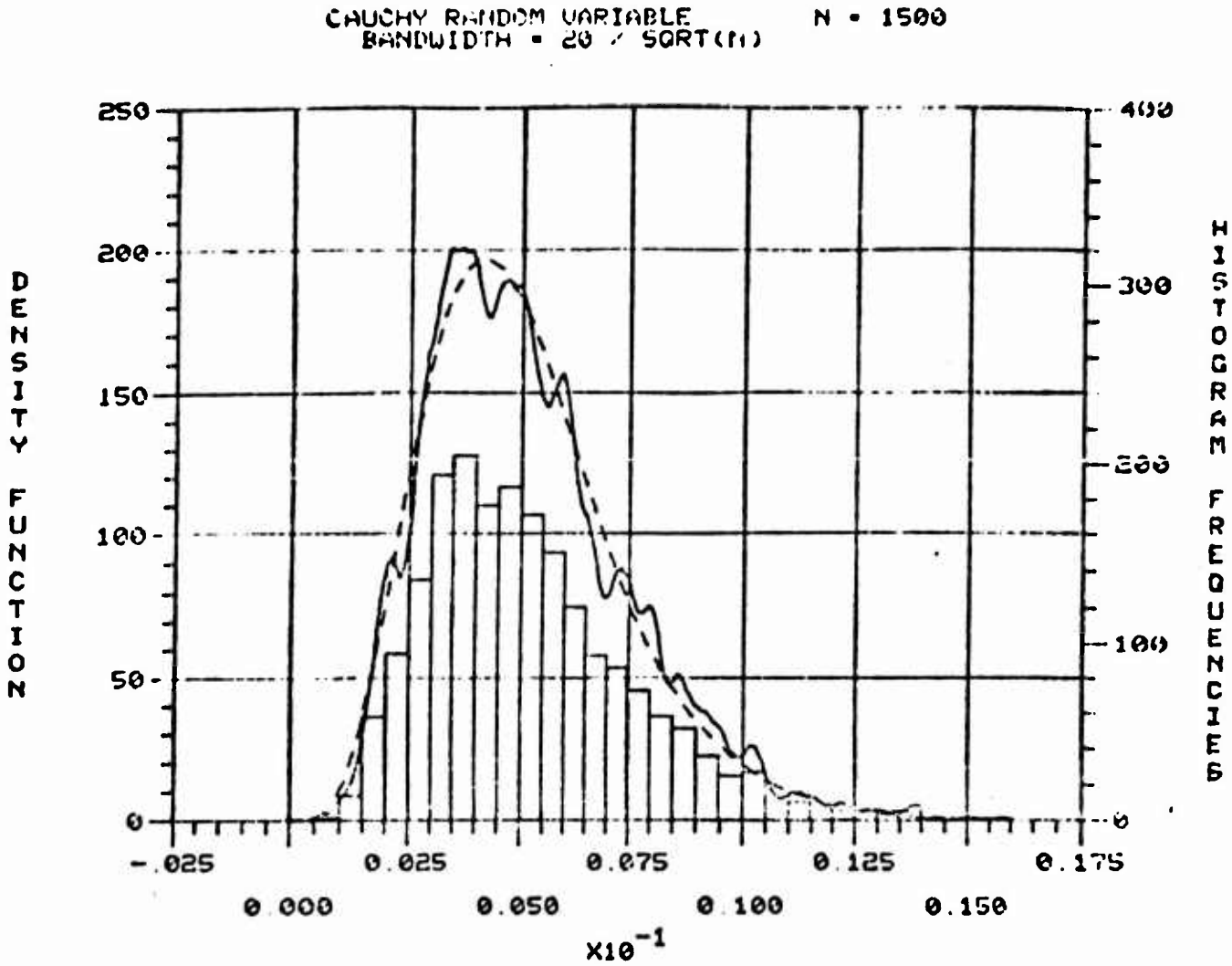

Figure 4.7. Distribution of the statistic $\beta(n)$ for a Cauchy random variable with n = 1500 and bandwidth $20 / \sqrt{n}$. The solid line shows the Rosenblatt empirical density function of the $\beta(n)$'s while the dashed line is a fitted Gamma density function with $\bar{K} = 5.385$ and $\bar{\theta} = .00095$.

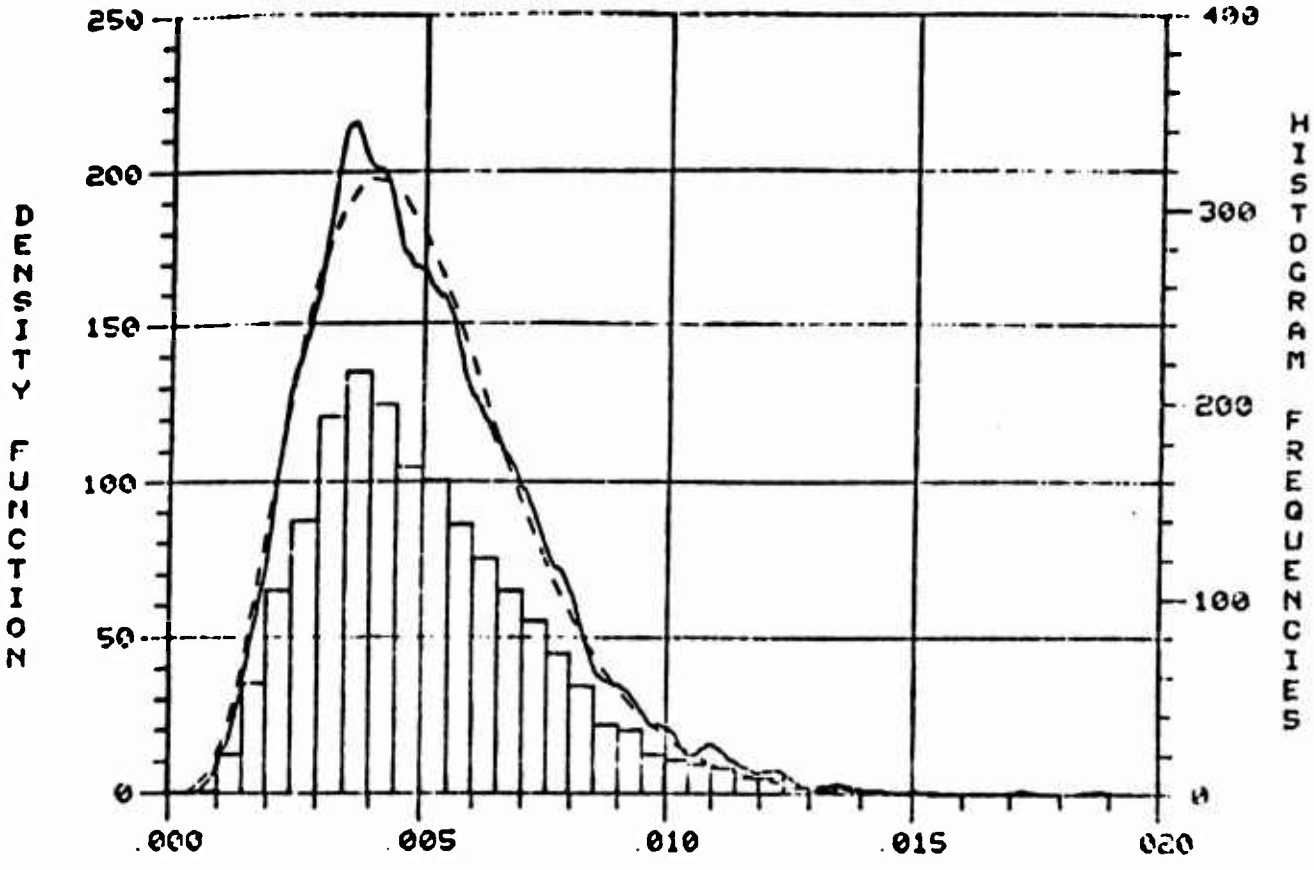

Figure 4.8. Distribution of the statistic $\beta(n)$ for a uniform random variable with $n = 1500$ and bandwidth $3 / \sqrt{n}$. The solid line shows the Rosenblatt empirical density function of the $\beta(n)$'s while the dashed line is a fitted Gamma density function with $K = 5.248$ and $\theta = .00096$.

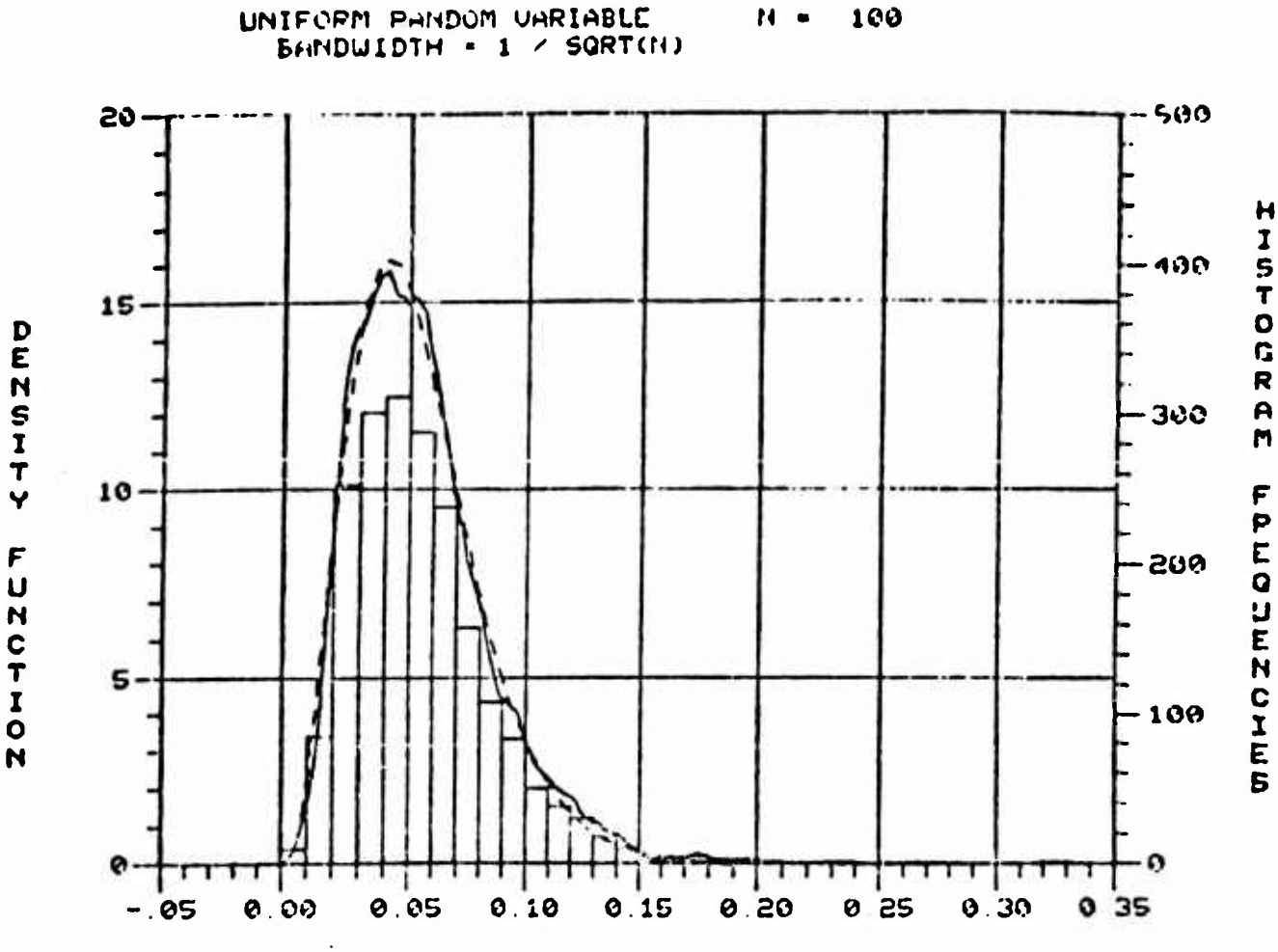

Figure 4.9. Distribution of the statistic $\beta(n)$ for a uniform random variable with n = 100 and bandwidth $1 / \sqrt{n}$. The solid line shows the Rosenblatt empirical density function of the $\beta(n)$'s while the dashed line is a fitted Gamma density function with $\hat{R} = 3.969$ and $\hat{\theta} = .01390$.

REFERENCES

[1] Bartlett, M.S., (1963). Statistical estimation of density functions. Sankhya, Ser. A25, p. 245-254.

[2] Bickel, P.J. and Rosenblatt, M., (1973). On some global measures of the deviations of density function estimates. The Annals of Mathematical Statistics, v. 1, p. 1071-1095.

[3] Liu, L.H., (1974). Empirical sampling investigation of a global measure of fit of probability density functions. M.S. Thesis, Naval Postgraduate School, Monterey.

[4] Rosenblatt, M., (1956). Remarks on some non-parametric estimates of a density function. The Annals of Mathematical Statistics, v. 27.

[5] Rosenblatt, M., (1971). Curve estimates. The Annals of Mathematical Statistics, v. 42.

[6] Shenton, L.R., and Bowman, K.O., (1973). Comments on the Gamma distribution and uses in rainfall data. Third Conference on Probability and Statistics in Atmospheric Science, AMS.

[7] Wegman, E.J., (1972). Non-parametric probability density estimtion: I. A summary of available methods. Technometrics, v. 14.